

Grassmannian Differential Limited Feedback for Interference Alignment

Omar El Ayach and Robert W. Heath, Jr.

The University of Texas at Austin

Department of Electrical & Computer Engineering

1 University Station C0806

Austin, TX, 78712-0240

Email: {omarayach, rheath}@mail.utexas.edu

Abstract

Channel state information (CSI) in the interference channel can be used to precode, align, and reduce the dimension of interference at the receivers, to achieve the channel's maximum multiplexing gain, through what is known as interference alignment. Most interference alignment algorithms require knowledge of all the interfering channels to compute the alignment precoders. CSI, considered available at the receivers, can be shared with the transmitters via limited feedback. When alignment is done by coding over frequency extensions in a single antenna system, the required CSI lies on the Grassmannian manifold and its structure can be exploited in feedback. Unfortunately, the number of channels to be shared grows with the square of the number of users, creating too much overhead with conventional feedback methods. This paper proposes Grassmannian differential feedback to reduce feedback overhead by exploiting both the channel's temporal correlation and Grassmannian structure. The performance of the proposed algorithm is characterized both analytically and numerically as a function of channel length, mobility, and the number of feedback bits. The main conclusions are that the proposed feedback strategy allows interference alignment to perform well over a wide range of Doppler spreads, and to approach perfect CSI performance in slowly varying channels. Numerical results highlight the trade-off between the frequency of feedback and the accuracy of individual feedback updates.

I. INTRODUCTION

Interference alignment (IA) is a precoding strategy that attempts to align interfering signals in time, frequency, or space. By aligning signals, IA reduces the dimension of the interference observed at each receiver and achieves the channel's maximum multiplexing gain. With alignment, network sum rate can increase linearly with the number of communicating users [1]. Various methods of calculating IA precoders have been proposed [2]–[9]. While [2]–[9] are algorithmically different, they share the common feature that channel state information (CSI) for the interference channels is required to compute the IA solution. In network architectures envisioned for IA, the channels would be known at a centralized processor that computes the IA precoders and distributes the solution to the transmitters, or each transmitter would have joint knowledge of all the CSI to compute the solution independently. In either case, a reasonable approach for satisfying the CSI requirement is to feedback quantized CSI through limited feedback, of which a detailed overview can be found in [10] and the references therein. A straightforward application of limited feedback, however, incurs significant overhead in IA networks. The reasons are that (i) feedback about every interfering link is required from every receiver and (ii) high resolution CSI is required to achieve alignment. As a result, low-overhead methods to transfer CSI must be designed to realize the data rates promised by IA.

Several approaches have been proposed to reduce feedback overhead in IA networks [11]–[15]. In [11], knowledge of channel gains is used to partition the network into optimally sized alignment sub-groups. In [12], IA is applied to partially connected interference channels. Exploiting the network structure as in [11], [12] limits the number of channels that must be fed back but does not provide better compression or higher quality CSI. To reduce the number of bits needed to feedback a channel response, [13] uses Grassmannian codebooks to quantize the channel coefficients in single-input single-output (SISO) systems with frequency extensions. The work in [14] extends that of [13] to systems with multiple antennas. Both [13] and [14] show that IA with limited feedback can achieve a system's degrees of freedom as long as feedback bits scale sufficiently with SNR. As the number of feedback bits increases, however, complexity increases and limited feedback becomes less practical as very large codebooks are undesirable. Analog feedback overcomes the problem of scaling complexity since the CSI distortion decreases naturally with SNR [15]. Analog feedback, however, is only optimal when the number of feedback channel uses matches the number of symbols being fed back [16]. Further, a main limitation of the feedback strategies in [13]–[15] is that they neglect the temporal correlation in the channel that can be exploited to further reduce overhead and improve CSI quality.

Differential and predictive feedback strategies have been proposed for compressing various forms of CSI in time-varying channels [17]–[24]. Simplified differential strategies, such as [17], have been considered in commercial wireless standards to feedback channel quality indicators. In [19] differential feedback of multiple-input multiple-output (MIMO) channel correlation matrices is proposed by leveraging the geometry of positive definite Hermitian matrices. The strategy in [19] was shown to work well when implemented on a wireless testbed [20]. In [21] the differential feedback of unitary precoders is considered in a single-user MIMO setting. The geometry of IA’s required CSI, however, neither fits the structure considered in [17], [19]–[21], nor the parametrization derived in [22]. Thus, the feedback strategies in [17], [19]–[22] cannot be used. In [18], tree-structured beamforming codebooks are constructed allowing receivers to feedback differentially encoded information about a local neighborhood in the feedback tree. In [23] differential feedback is applied to beamforming vectors that are gradually rotated on the Grassmann manifold using a Householder matrix to improve quantization accuracy. In [24] Grassmannian beamforming vectors are fed back and reconstructed using predictive vector quantization tools, leveraging manifold concepts such as tangent vectors and geodesic curves. Unlike the single-user system considered in [23], however, IA systems are sensitive to leakage interference, and require more feedback than the broadcast channel [18], [24]. As a result, the application of [18], [23], [24] to IA networks is non-trivial and it is not clear whether they can achieve sufficient performance with practical codebook sizes.

In this paper, we propose a limited feedback strategy based on Grassmannian differential feedback to fulfill the CSI requirement in wideband single antenna systems using IA over frequency extensions. Frequency extensions provide a practical source of dimensionality for IA precoding in SISO interference channels and play an integral role in achieving the maximum multiplexing gain in both single and multiple antenna systems [1]. Although we present results for single antenna systems, the feedback strategy can be generalized to systems with multiple antennas in a manner similar to [14]. The proposed strategy tracks the slow evolution of the normalized channel impulse responses on the Grassmannian manifold. At each feedback update, a quantized tangent vector relating consecutive channel realizations is used at the transmitter to reconstruct the channel. Since the tangent space geometry changes at each iteration, and varying Doppler spreads lead to different channel dynamics, we construct quantization codebooks that adapt to the channel’s geometry and dynamics. Unlike [24], no prediction is performed and differential feedback is used in the new context of SISO interference channels. We show that the performance of the proposed differential strategy cannot be improved upon by a simple yet commonly used class of linear or geodesic predictors. Thus, the development of more sophisticated prediction algorithms that can further improve CSI quality remains a promising venue for future work.

We derive an approximation for the average distortion achieved by the proposed algorithm and show that it is accurate for a wide range of channel and feedback parameters. Unlike the analysis of the Householder strategy in [23] and the predictive strategy in [24], which neglect either the quantization process or the channel's Doppler spread, the derived performance characterization captures the effects of both fading speed and codebook size. Simulation results show that the proposed algorithm outperforms memoryless quantization and other competitive feedback strategies for temporally correlated channels [24]. Through simulation, we highlight also the trade-off between feedback refresh rate and codebook size [25]. We show that, from a CSI distortion perspective, frequent updates using smaller codebooks may often be preferred to less-frequent updates with a large codebook. Finally, while scaling codebook size is still required to preserve multiplexing gain, we show that IA with the proposed feedback strategy performs well in the range of medium to high SNR with a relatively small number of feedback bits.

This paper is organized as follows. Section II introduces the SISO wideband interference channel model with orthogonal frequency division multiplexing (OFDM). Section III reviews the concept of interference alignment over frequency extended channels and highlights the degradation due to imperfect channel knowledge. Section IV presents the proposed feedback framework while Section V elaborates on its main design parameters. Sections VI and VII present analytical and numerical results respectively on the performance of the Grassmannian differential feedback, as well as the performance of IA when the proposed strategy is used for CSI feedback. We conclude with Section VIII.

We summarize the notation employed in this paper: \mathbf{A} is a matrix; \mathbf{a} is a vector; and a is a scalar; $(\cdot)^*$ denotes the conjugate transpose; $\mathbf{A} \circ \mathbf{B}$ is the Hadamard product; $\|\mathbf{A}\|_F$ is the Frobenius norm of \mathbf{A} and $\text{trace}(\mathbf{A})$ is its trace; $\|\mathbf{a}\|$ is the 2-norm of \mathbf{a} ; $|a|$ is the absolute value of a ; \mathbf{I}_N is the $N \times N$ identity matrix; $\mathbf{0}_N$ is the N -dimensional zero vector; \mathcal{F}_N is the N -point discrete Fourier transform (DFT) matrix; $\text{diag}(\mathbf{a})$ is the diagonal matrix obtained by putting the elements of \mathbf{a} on its diagonal; $[\mathbf{a}, \mathbf{b}]$ is a horizontal concatenation of vectors \mathbf{a} and \mathbf{b} ; \mathbb{C}^N is the N -dimensional complex space; $\mathcal{CN}(\mathbf{a}, \mathbf{A})$ is a complex Gaussian random vector with mean \mathbf{a} and covariance matrix \mathbf{A} . Expectation is denoted by $\mathbb{E}[\cdot]$.

II. SYSTEM MODEL

Consider a frequency selective SISO interference channel with K communicating user pairs. Each user k communicates desired data to its paired receiver k and interferes with all other receivers $\ell \neq k$. The wideband channel between transmitter ℓ and receiver k at time t is given by the L -tap channel impulse response vector $\mathbf{h}_{k,\ell}[t] = [h_{k,\ell}[t, 0]^*, h_{k,\ell}[t, 1]^*, \dots, h_{k,\ell}[t, L-1]^*]^* \forall k, \ell \in \{1, \dots, K\}$. Channel impulse responses are drawn independently across k and ℓ from a continuous distribution and have covariance

matrix $\mathbb{E} [\mathbf{h}_{k,\ell}[t]\mathbf{h}_{k,\ell}[t]^*] = \mathbf{R}_{\mathbf{h}_{k,\ell}}, \forall k, \ell \in \{1, \dots, K\}$ and $\forall t$. Letting $\rho_{k,\ell}$ be the average attenuation of user ℓ 's signal at user k 's receiver, the channel response satisfies $\text{trace}(\mathbf{R}_{\mathbf{h}_{k,\ell}}) = \rho_{k,\ell}$. In the case of uncorrelated channel taps, $\mathbf{R}_{\mathbf{h}_{k,\ell}} = \text{diag}(\mathbf{p}_{k,\ell})$, where $\mathbf{p}_{k,\ell}$ is the $L \times 1$ vector representing the channel's power delay profile.

Using orthogonal frequency division multiplexing (OFDM), the transmitters transform the observed frequency selective channels into a set of N_{sc} non-interfering narrowband subcarriers. Stacking each received OFDM symbol in a vector, the input-output relationship can be written in matrix form as

$$\mathbf{y}_k[t] = \mathbf{H}_{k,k}[t]\mathbf{x}_k[t] + \sum_{\ell \neq k} \mathbf{H}_{k,\ell}[t]\mathbf{x}_\ell[t] + \mathbf{v}_k[t], \quad (1)$$

where $\mathbf{x}_k[t]$ is the OFDM symbol sent by user k at time t with the average power constraint $\mathbb{E} [\|\mathbf{x}_k[t]\|^2] = N_{\text{sc}}P$, the $N_{\text{sc}} \times N_{\text{sc}}$ matrix $\mathbf{H}_{k,\ell}[t] = \text{diag} \left(\mathcal{F}_{N_{\text{sc}}} \left[\mathbf{h}_{k,\ell}^*[t], 0_{N_{\text{sc}}-L} \right]^* \right)$ represents the channel frequency response between transmitter ℓ and receiver k at time t , and $\mathbf{v}_k[t]$ is the i.i.d. $\mathcal{CN}(0, \sigma_n^2 \mathbf{I}_{N_{\text{sc}}})$ thermal noise observed by user k . The system model assumes perfect time and frequency synchronization, and a cyclic prefix that is long enough to accommodate the impulse response of all channels as well as the potential propagation delay between users.

Further, the relationship in (1) implicitly assumes that the L -tap channels $\mathbf{h}_{k,\ell}[t]$ seen by the t^{th} OFDM symbol remain constant over the OFDM symbol time and pilots can be used to estimate them at the receiver. The channels over consecutive OFDM symbols, however, are considered to be slowly varying. Let f_D be the Doppler frequency of the channel and T_s be the OFDM symbol time. We assume channels are temporally correlated according to the model proposed by Clarke [26], i.e $\mathbb{E} \left[\rho_{k,\ell}^{-1} \left| \mathbf{h}_{k,\ell}^*[t-m]\mathbf{h}_{k,\ell}[t] \right| \right] = J_0(2\pi f_D T_s m)$ where J_0 is the 0-th order Bessel function of the first kind. A main underlying assumption, however, is that while the channels vary slowly over time, their length L remains fixed. Since channel response length, is likely to vary slower than the channel response itself, and since it is constrained to be a small integer, the channel length can be fed back to the transmitter if necessary at a small extra overhead cost. In general, the proposed feedback algorithm can be applied to all correlated channels, however, further details on the mathematical structure of the temporally correlated channel are given whenever a result is model-dependent.

Throughout the remainder of this paper we make the following two system-level assumptions:

- We assume that the channels $\mathbf{h}_{k,\ell}[t] \forall \ell$ are known to receiver k . This implicitly assumes that the set of K transmitters coordinate (or orthogonalize) their pilot transmissions allowing receivers to estimate all incoming channels. We make a further simplification by assuming receiver CSI is perfect.

- We assume each receiver has an error-free logical feedback link to all transmitters. The feedback link from receiver k to transmitter ℓ may either be a direct link, or a link through other nodes such as transmitter k and/or a centralized processor. Effectively, this allows obtaining global CSI at at least one node to enable the calculation of IA precoders which, if needed, can be forwarded to the other transmitters.

While the importance of transmitter channel knowledge makes the two assumptions rather standard in the literature on precoding for the interference channel [1], [2], [6]–[8], [11]–[15], or interference coordination in general [27], they are by no means trivial. Coordinating training phases, feedback phases, and centralized processing constitutes significant complexity in both cellular and ad-hoc systems. We refer the interested reader to [27] and the references therein for a detailed discussion of the limitations, the benefits, and the possible solutions that enable coordination in the case of cellular systems.

III. INTERFERENCE ALIGNMENT IN FREQUENCY

In this section we review the concept of IA over frequency extensions when perfect channel state information is available at the transmitter \mathbf{W} then summarize the effect of imperfect transmitter channel knowledge on the performance of IA.

A. IA with Perfect CSI at the Transmitter

IA for the SISO interference channel can achieve the maximum degrees of freedom when coding over infinite channel extensions [1]. Using IA over N frequency extensions, each transmitter k at time t sends $d_k < N$ symbols, $x_k^m[t]$, along the $N \times 1$ precoding vectors $\mathbf{f}_k^m[t]$. As a result, the transmitted symbol is

$$\mathbf{x}_k[t] = \sum_{m=1}^{d_k} \mathbf{f}_k^m[t] x_k^m[t], \quad (2)$$

where $\|\mathbf{f}_k^m[t]\|_2 = 1$ and $\mathbb{E}[|x_k^m[t]|^2] = NP/d_k$, such that the power in each N subcarriers is NP [28]. The transmit directions $\mathbf{f}_k^m[t]$ are calculated such that the interference from $K - 1$ users is aligned at all receivers, leaving at least d_k interference free dimensions for the desired signal. The number of transmitted symbols d_k can be chosen according to the original strategy in [1] or the improved method in [7], both of which asymptotically achieve the maximum degrees of freedom, i.e. $\lim_{N \rightarrow \infty} \frac{1}{N} \sum d_k = K/2$.

Note that users need not code over all N_{sc} subcarriers since algorithm complexity increases significantly with the number of subcarriers whereas the achieved multiplexing gain may increase very slowly with N_{sc} as demonstrated by the IA solutions in [1]. Users could potentially treat any N -subset of subcarriers

as an independent coding group and thus have $\lfloor \frac{N_{sc}}{N} \rfloor$ parallel alignment groups in an OFDM symbol. In this case (2) applies to each group of subcarriers, and one can write $\lfloor \frac{N_{sc}}{N} \rfloor$ relationships for each OFDM symbol. For simplicity of exposition, in the remainder of this section we assume $\lfloor \frac{N_{sc}}{N} \rfloor = 1$.

To quantify concisely the degradation due to imperfect CSI, we consider a zero-forcing receiver; other receiver designs can be used. At the output of the linear receivers $\mathbf{w}_k^m[t]$, the received signal is

$$\begin{aligned} \mathbf{w}_k^m[t]^* \mathbf{y}_k[t] = & \mathbf{w}_k^m[t]^* \mathbf{H}_{k,k}[t] \mathbf{f}_k^m[t] x_k^m[t] + \sum_{\ell \neq m} \mathbf{w}_k^m[t]^* \mathbf{H}_{k,k}[t] \mathbf{f}_k^\ell[t] x_k^\ell[t] \\ & + \sum_{i \neq k} \sum_{\ell=1}^{d_i} \mathbf{w}_k^m[t]^* \mathbf{H}_{k,i}[t] \mathbf{f}_i^\ell[t] x_i^\ell[t] + \mathbf{w}_k^m[t]^* \mathbf{v}_k[t], \end{aligned} \quad (3)$$

for $m \in \{1, \dots, d_k\}$ and $k \in \{1, \dots, K\}$, where $\|\mathbf{w}_k^m[t]\|^2 = 1$. Assuming a zero-forcing receiver, the conditions for perfect interference alignment are

$$\mathbf{w}_k^m[t]^* \mathbf{H}_{k,k}[t] \mathbf{f}_k^\ell[t] = 0 \quad \forall k, \ell \neq m \quad (4)$$

$$\mathbf{w}_k^m[t]^* \mathbf{H}_{k,i}[t] \mathbf{f}_i^\ell[t] = 0 \quad \forall i \neq k, \text{ and } \forall m, \ell \quad (5)$$

$$|\mathbf{w}_k^m[t]^* \mathbf{H}_{k,k}[t] \mathbf{f}_k^m[t]| \geq c > 0 \quad \forall k, m \quad (6)$$

where alignment is achieved by (4) and (5), while (6) ensures the decodability of the d_k desired streams.

The sum rate achieved by the linear zero-forcing receiver [29], assuming Gaussian input signals, is

$$\bar{R}_{\text{sum}} = \sum_{k=1}^K \sum_{m=1}^{d_k} \frac{1}{N_{sc}} \log_2 \left(1 + \frac{\frac{N_{sc}P}{d_k} |\mathbf{w}_k^m[t]^* \mathbf{H}_{k,k}[t] \mathbf{f}_k^m[t]|^2}{\mathcal{I}_{k,m}^1[t] + \mathcal{I}_{k,m}^2[t] + \sigma_n^2} \right), \quad (7)$$

where $\mathcal{I}_{k,m}^1[t]$ is self-interference from other transmit streams and $\mathcal{I}_{k,m}^2[t]$ is the interference from other users. The two interference terms are

$$\begin{aligned} \mathcal{I}_{k,m}^1[t] &= \sum_{\ell \neq m} \frac{N_{sc}P}{d_k} \left| \mathbf{w}_k^m[t]^* \mathbf{H}_{k,k}[t] \mathbf{f}_k^\ell[t] \right|^2, \\ \mathcal{I}_{k,m}^2[t] &= \sum_{i \neq k} \sum_{\ell=1}^{d_i} \frac{N_{sc}P}{d_i} \left| \mathbf{w}_k^m[t]^* \mathbf{H}_{k,i}[t] \mathbf{f}_i^\ell[t] \right|^2. \end{aligned} \quad (8)$$

If perfect channel knowledge is available, and the number of symbols d_k are feasible, equations (4)-(6) can be satisfied with probability one and thus $\mathcal{I}_{k,m}^1 = \mathcal{I}_{k,m}^2 = 0$. This gives

$$\lim_{P \rightarrow \infty} \frac{\bar{R}_{\text{sum}}}{\log_2 P} = \lim_{P \rightarrow \infty} \frac{\sum_{k,m} \frac{1}{N_{sc}} \log_2 \left(1 + \frac{\frac{N_{sc}P}{d_k} |\mathbf{w}_k^m[t]^* \mathbf{H}_{k,k}[t] \mathbf{f}_k^m[t]|^2}{\sigma_n^2} \right)}{\log_2 P} = \frac{1}{N_{sc}} \sum_k d_k \xrightarrow{N_{sc} \rightarrow \infty} \frac{K}{2}, \quad (9)$$

which confirms the fact that IA achieves the $K/2$ degrees of freedom of the K user interference channel. While the achievability proof in [1] assumed independent fading across subcarriers, [13] has claimed that fading on each subcarrier need not be independent provided that the channel impulse response is long enough.

B. The Effect of Limited Feedback

In practical systems, receivers cannot feedback the channels $\mathbf{h}_{k,i}[t]$ with infinite precision, thus only a distorted version of $\mathbf{h}_{k,i}[t]$ is available when calculating IA precoders. The distortion caused by limited feedback implies that interference will not be perfectly aligned. The power of residual misaligned interference increases along with transmit power and causes the SINR in the desired signal subspace to saturate, which limits sum rate at high SNR [29]. As a result, low overhead feedback strategies must be designed to improve CSI accuracy and allow good IA performance.

To establish the impact of limited feedback on IA sum rate, and to develop insights into the structure of a good quantizer, we summarize the calculations in [13], [15]. In [15], it is shown that if the transmitters use the estimated or quantized channels $\hat{\mathbf{H}}_{k,i}$, $\forall k, i \in \{1, 2, \dots, k\}$ to calculate IA precoders, then the mean loss in sum rate can be upper bounded by

$$\Delta \bar{R}_{\text{sum}} \leq \sum_{k,m} \frac{1}{N_{\text{sc}}} \log_2 \left(1 + \frac{\mathbb{E}_{\mathbf{H}} [\mathcal{I}_{k,m}^1 + \mathcal{I}_{k,m}^2]}{\sigma_n^2} \right). \quad (10)$$

Using the results from [13, Section IV-B], the components of the sum leakage interference, $\mathcal{I}_{k,m}^1 + \mathcal{I}_{k,m}^2$, can be bounded by

$$\frac{N_{\text{sc}} P}{d_i} \left| \hat{\mathbf{w}}_k^m[t]^* \mathbf{H}_{k,i}[t] \hat{\mathbf{f}}_i^\ell[t] \right|^2 \leq \frac{N_{\text{sc}} P}{d_i} \|\hat{\mathbf{w}}_k^m[t] \circ \hat{\mathbf{f}}_i^\ell[t]\|^2 \|\mathbf{h}_{k,i}[t]\|^2 \left(1 - \left| \frac{\mathbf{h}_{k,i}[t]^* \hat{\mathbf{h}}_{k,i}[t]}{\|\mathbf{h}_{k,i}[t]\| \|\hat{\mathbf{h}}_{k,i}[t]\|} \right|^2 \right), \quad (11)$$

where $\hat{\mathbf{f}}_i^\ell[t]$ and $\hat{\mathbf{w}}_k^m[t]$ are the precoders and combiners calculated using imperfect CSI. The bound in (11) relates the power of leakage interference to the quality of the quantized CSI and suggests mathematical structure that can be exploited by the quantizer. The last term in (11) indicates that only knowledge of the normalized channels $\frac{\mathbf{h}_{k,i}[t]}{\|\mathbf{h}_{k,i}[t]\|}$ is necessary for IA, further the CSI is phase invariant since $\left| \frac{\mathbf{h}_{k,i}[t]^* \hat{\mathbf{h}}_{k,i}[t]}{\|\mathbf{h}_{k,i}[t]\| \|\hat{\mathbf{h}}_{k,i}[t]\|} \right|^2 = \left| \frac{e^{j\theta} \mathbf{h}_{k,i}[t]^* \hat{\mathbf{h}}_{k,i}[t]}{\|e^{j\theta} \mathbf{h}_{k,i}[t]\| \|\hat{\mathbf{h}}_{k,i}[t]\|} \right|^2$. This is not surprising since, if an interfering channel is scaled and rotated by $\alpha \in \mathbb{C}$, the interference subspace remains unchanged, i.e., $\text{diag}(\mathcal{F}_{N_{\text{sc}}} \alpha \mathbf{h}_{k,i}[t]) \mathbf{f}_i^m[t] = \alpha \text{diag}(\mathcal{F}_{N_{\text{sc}}} \mathbf{h}_{k,i}[t]) \mathbf{f}_i^m[t]$. As a result, the CSI needed for IA with frequency extensions evolves on the manifold of L -dimensional unit norm, rotationally invariant vectors otherwise known as the Grassmannian

manifold $\mathcal{G}_{L,1}$ [30]–[33]. Define the chordal distance between two points on the Grassmannian manifold, \mathbf{x}_1 and \mathbf{x}_2 , as $d(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{1 - |\mathbf{x}_1^* \mathbf{x}_2|^2}$. The bound in (11) can be rewritten as

$$\frac{N_{sc}P}{d_i} \left| \widehat{\mathbf{w}}_k^m[t]^* \mathbf{H}_{k,i}[t] \widehat{\mathbf{f}}_i^\ell[t] \right|^2 \leq \frac{N_{sc}P}{d_i} \|\widehat{\mathbf{w}}_k^m[t] \circ \widehat{\mathbf{f}}_i^\ell[t]\|^2 \|\mathbf{h}_{k,i}[t]\|^2 d(\mathbf{g}_{k,i}[t], \widehat{\mathbf{g}}_{k,i}[t])^2, \quad (12)$$

where $\mathbf{g}_{k,i}[t] \triangleq \frac{\mathbf{h}_{k,i}[t]}{\|\mathbf{h}_{k,i}[t]\|}$ and $\widehat{\mathbf{g}}_{k,i}[t] \triangleq \frac{\widehat{\mathbf{h}}_{k,i}[t]}{\|\widehat{\mathbf{h}}_{k,i}[t]\|}$, showing that leakage interference can be bounded using the chordal distance between the true and quantized channels [34].

IV. GRASSMANNIAN DIFFERENTIAL FEEDBACK

In this section, we present the proposed feedback algorithm. Under some conditions, we show that differential quantization, a special case of predictive vector quantization, cannot be improved upon using a simple yet common class of predictors.

A. Differential Feedback Framework

We propose to differentially encode the normalized channel response vectors by exploiting both their Grassmannian structure and temporal correlation. The proposed algorithm encodes CSI increments using a tangent vector which defines the geodesic path between consecutive channel realizations. Our feedback framework uses tools presented in [34], which were simplified and first used for MISO broadcast channel feedback in [24]. We consider the separate quantization of the slowly varying normalized channel vectors $\mathbf{g}_{k,i}[t] \triangleq \frac{\mathbf{h}_{k,i}[t]}{\|\mathbf{h}_{k,i}[t]\|} \forall t \geq 0$; joint quantization may offer higher performance at the expense of additional complexity [35]. Therefore, we restrict our attention to one of the channels and drop the user subscripts.

The smooth structure of the manifold allows each two points to be related using the concept of a geodesic path as shown in Fig. 1. The geodesic curve describing the path between $\mathbf{g}[t-1]$ and $\mathbf{g}[t]$ is given by

$$G(\mathbf{g}[t-1], \mathbf{e}[t], \ell) = \mathbf{g}[t-1] \cos(\|\mathbf{e}[t]\| \ell) + \frac{\mathbf{e}[t]}{\|\mathbf{e}[t]\|} \sin(\|\mathbf{e}[t]\| \ell). \quad (13)$$

where the vector $\mathbf{e}[t]$, also shown in Fig. 1, is known as the tangent vector. The tangent vector can be viewed as a length-preserving unwrapping of the geodesic path onto the tangent space at $\mathbf{g}[t-1]$. The tangent vector is given by

$$\mathbf{e}[t] = \tan^{-1} \left(\frac{d[t]}{\rho[t]} \right) \frac{\mathbf{g}[t]/\rho[t] - \mathbf{g}[t-1]}{d[t]/\rho[t]}, \quad (14)$$

where $\rho[t] = \mathbf{g}[t-1]^* \mathbf{g}[t]$ and $d[t] = \sqrt{1 - |\rho[t]|^2}$ [34]. From (13), it can be verified that $G(\mathbf{g}[t-1], \mathbf{g}[t], 0) = \mathbf{g}[t-1]$, $G(\mathbf{g}[t-1], \mathbf{g}[t], 1) = \mathbf{g}[t]$ and that $\|G(\mathbf{g}[t-1], \mathbf{g}[t], \ell)\| = 1, \forall \ell$ since the tangent vector $\mathbf{e}[t]$ is orthogonal to the base vector $\mathbf{g}[t-1]$.

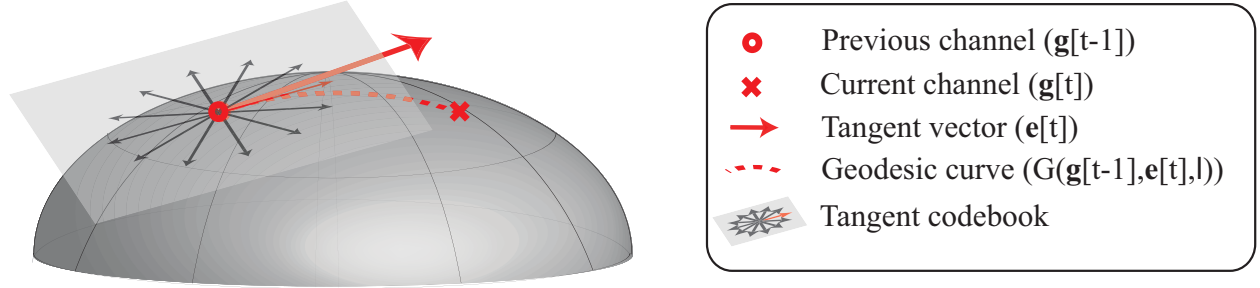


Fig. 1. A 3D visualization of the Grassmann manifold and the tangent vector quantization process. The quantized tangent which is selected from the tangent codebook is also marked in red.

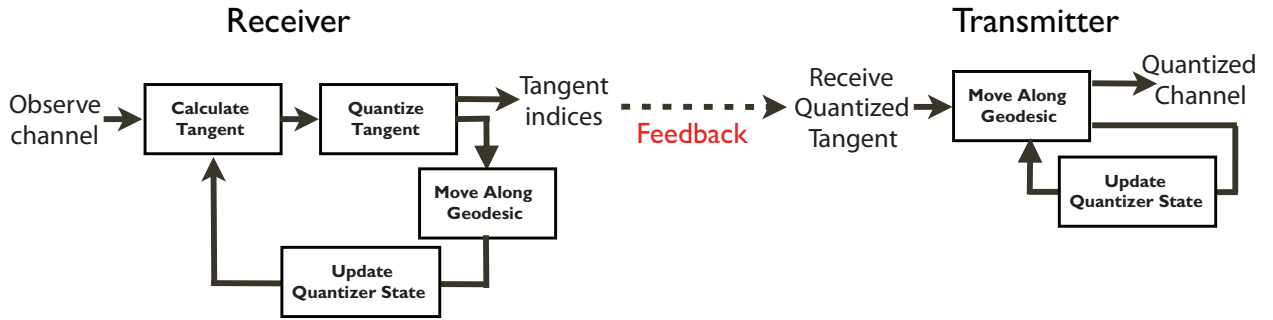


Fig. 2. Block diagram of Grassmannian differential quantization and feedback at the transmitter and receiver.

The tangent vector and geodesic path, which together relate two points on the manifold, can be used to build a differential feedback framework to track the evolution of CSI [24]. Given the previous channel realization and a tangent vector received by feedback, the transmitter can reconstruct the current channel by applying (13). The operation of the Grassmannian differential feedback algorithm at both the transmitter and receiver is given in the block diagram of Fig. 2. For clarity, the pseudo code used to encode the channel evolution over time is given in Algorithm 1. At each new channel realization, the receiver estimates the normalized L -tap channel $\mathbf{g}[t]$. Perfect channel estimation is assumed here to decouple the quantization error from the estimation error. Given the current observation, and the quantized channel in the previous iteration, $\hat{\mathbf{g}}[t-1]$, the receiver calculates the tangent vector, $\mathbf{e}[t]$, from $\hat{\mathbf{g}}[t-1]$ to $\mathbf{g}[t]$. The receiver then quantizes the tangent vector to produce a quantized tangent, $\hat{\mathbf{e}}[t]$. The details of the quantization method are given in depth in Section V. This quantized tangent is then fed back to the transmitter over a delay and error free feedback link. Equipped with the previous quantized channel, $\hat{\mathbf{g}}[t-1]$, and the quantized error vector $\hat{\mathbf{e}}[t]$, *both transmitter and receiver* can now calculate the new quantized channel $\hat{\mathbf{g}}[t]$. This vector $\hat{\mathbf{g}}[t]$ will later serve as the base point in the next iteration. The

Algorithm 1 Receiver

```

1: Input:  $\hat{\mathbf{g}}[t-1]$ 
2: for all  $t = 1, 2, \dots$  do
3:   Estimate the channel  $\mathbf{g}[t]$ 
4:   Calculate tangent between  $\hat{\mathbf{g}}[t-1]$  and  $\mathbf{g}[t]$  using (14)
5:   Quantize and feedback the quantized tangent vector  $\hat{\mathbf{e}}[t]$ 
6:   Reconstruct the quantized channel  $\hat{\mathbf{g}}[t] = G(\hat{\mathbf{g}}[t-1], \hat{\mathbf{e}}[t], 1)$ 
7: end for

```

Algorithm 2 Transmitter

```

1: Input:  $\hat{\mathbf{g}}[t-1]$ 
2: for all  $t = 1, 2, \dots$  do
3:   Receive the feedback tangent vector  $\hat{\mathbf{e}}[t]$ 
4:   Reconstruct the quantized channel  $\hat{\mathbf{g}}[t] = G(\hat{\mathbf{g}}[t-1], \hat{\mathbf{e}}[t], 1)$ 
5: end for
6: Output:  $\hat{\mathbf{g}}[t]$ 

```

algorithm runs in parallel for all channels $\mathbf{g}_{k,i}[t]$, $\forall k, i$ and the transmitters use the estimated CSI vectors $\hat{\mathbf{g}}_{k,i}[t]$, $\forall k, \ell$ to do interference alignment.

B. Comparing Differential vs. Predictive Quantization

The feedback framework of Section IV-A is a special case of predictive vector quantization, the main difference being that the proposed strategy does not attempt to further improve CSI by *predicting future channel realizations*. In this section, we begin by formalizing the main difference between predictive and differential feedback. We then show that, for a class of channels, commonly used Grassmannian predictors such as those in [23], [24], yield no performance enhancement over the differential strategy presented. We note, however, that more sophisticated prediction functions may in fact further improve performance. Thus, the development of more intelligent Grassmannian predictors is a promising area for future work.

1) *Predictive Quantization:* The main idea behind differential feedback is to send the transmitter information that allows it to move from the *old quantized channel* to the *new channel realization*. Thus, a tangent vector $\mathbf{e}[t]$ is calculated between the base point $\hat{\mathbf{g}}[t-1]$ and $\mathbf{g}[t]$, and is then quantized to give $\hat{\mathbf{e}}[t]$. Instead of starting the tangent vector at the point $\hat{\mathbf{g}}[t-1]$, general predictive quantization calculates the tangent starting at *another* point $\tilde{\mathbf{g}}[t]$, known as the *predicted channel vector*. The premise is that if $\tilde{\mathbf{g}}[t]$ is “closer” to $\mathbf{g}[t]$ than $\hat{\mathbf{g}}[t-1]$ (the base point used in differential feedback), quantization accuracy can be further improved. The main difficulty now lies in constructing prediction functions that actually

yield closer or better estimates of $\mathbf{g}[t]$.

Unfortunately, work on prediction on the Grassmann manifold is rather limited. In this paper, we consider a class of linear predictors used in [23], [24]. Though the predictors in [23], [24] are presented differently, the predicted channel vectors can be written in a unified manner as

$$\tilde{\mathbf{g}}[t] = G(\hat{\mathbf{g}}[t-1], \mathbf{e}_p[t], \ell_p) = \hat{\mathbf{g}}[t-1] \cos(\|\mathbf{e}_p[t]\| \ell_p) + \frac{\mathbf{e}_p[t]}{\|\mathbf{e}_p[t]\|} \sin(\|\mathbf{e}_p[t]\| \ell_p), \quad (15)$$

where $\mathbf{e}_p[t]$ is the *predicted tangent vector* and ℓ_p is a predicted step size. Since the norm of $\mathbf{e}_p[t]$ can absorb any prediction step size ℓ_p , we enforce $\ell_p = 1$ without loss of generality.

The challenge now is to use commonly known information, such as $\hat{\mathbf{g}}[t'] \forall 0 \leq t' \leq t-1$,¹ to find the best predicted tangent vector $\mathbf{e}_p^{\text{opt}}[t]$, i.e., the one that maximizes $\mathbb{E} \left[|\tilde{\mathbf{g}}[t]^* \mathbf{g}[t]|^2 \right]$ and makes $\tilde{\mathbf{g}}[t]$ the *best estimate* of $\mathbf{g}[t]$. Thus, the prediction problem can be summarized as solving

$$\begin{aligned} \mathbf{e}_p^{\text{opt}}[t] &= \arg \max_{\mathbf{e}_p[t]} \mathbb{E} \left[|\tilde{\mathbf{g}}[t]^* \mathbf{g}[t]|^2 \right] \\ &\stackrel{(a)}{=} \arg \max_{\mathbf{e}_p[t]} \mathbb{E} \left[\left| \left(\cos(\|\mathbf{e}_p[t]\|) \hat{\mathbf{g}}[t-1] + \sin(\|\mathbf{e}_p[t]\|) \frac{\mathbf{e}_p[t]}{\|\mathbf{e}_p[t]\|} \right)^* \mathbf{h}[t] \right|^2 \right], \end{aligned} \quad (16)$$

where (a) is by replacing $\tilde{\mathbf{g}}[t]$ by its definition in (15) and by noticing that maximizing $\mathbb{E} \left[|\tilde{\mathbf{g}}[t]^* \mathbf{g}[t]|^2 \right]$ is equivalent to maximizing $\mathbb{E} \left[|\tilde{\mathbf{g}}[t]^* \mathbf{h}[t]|^2 \right]$. In what follows, we show that for a class of temporally correlated channels, such a geodesic predictor yields no performance enhancement over the differential feedback strategy presented. Therefore, to ensure that prediction does in fact improve performance, more sophisticated predictors that better exploit the channel's Doppler spectrum must be derived. The development of such predictors is a promising topic for future work.

2) Performance Comparison & Prediction Gains: For the analysis in this section, we make the following assumptions on the channel's temporal structure.

Assumption 1: We assume that the channels $\mathbf{h}[t]$ are temporally correlated according to a first order autoregressive model [36], and are thus given by $\mathbf{h}[t] = \eta_f \mathbf{h}[t-1] + \sqrt{1-\eta_f^2} \mathbf{z}[t]$. For simplicity, we further assume that $\mathbf{h}[t] \sim \mathcal{CN}(0, \frac{1}{L} \mathbf{I}_L)$ which in turn implies that $\mathbf{z}[t] \sim \mathcal{CN}(0, \frac{1}{L} \mathbf{I}_L)$.

While we present our analysis for channels satisfying Assumption 1, similar results can be shown for other temporally correlated channel models in which changes in the channel are zero-mean and independent of the previous channel realizations.

¹In fact [23], [24] make the simplifying assumption that $\hat{\mathbf{g}}[t'] \forall 0 \leq t' \leq t-1$ is available during predictor design. We relax this assumption, by only allowing access to $\hat{\mathbf{g}}[t'] \forall 0 \leq t' \leq t-1$, at the cost of adding a milder assumption on the quantization error in $\hat{\mathbf{g}}[t]$. Results hold, and derivations are simplified, if we revert to the assumptions of [23], [24].

Assumption 2: We assume that the quantization error introduced by our algorithm is isotropically distributed, i.e., the direction of $\mathbf{g}[t] - \hat{\mathbf{g}}[t]$ is uniformly distributed, and that the error direction and magnitude are independent.

For a channel following a first order autoregressive model, the predictor's objective function, $\mathbb{E} [|\tilde{\mathbf{g}}[t]^* \mathbf{h}[t]|^2]$, can be written as

$$\begin{aligned} \mathbb{E} [|\tilde{\mathbf{g}}[t]^* \mathbf{h}[t]|^2] &= \mathbb{E} \left[\left| \left(\cos(\|\mathbf{e}_p[t]\|) \hat{\mathbf{g}}[t-1] + \sin(\|\mathbf{e}_p[t]\|) \frac{\mathbf{e}_p[t]}{\|\mathbf{e}_p[t]\|} \right)^* \left(\eta_f \mathbf{h}[t-1] + \sqrt{1-\eta_f^2} \mathbf{z}[t] \right) \right|^2 \right] \\ &= \mathbb{E} \left[\left| \eta_f (\hat{\mathbf{g}}[t-1], \mathbf{e}_p[t]) \boldsymbol{\beta}^* \mathbf{h}[t-1] + \bar{\eta}_f \boldsymbol{\beta}^* \mathbf{x}[t] \right|^2 \right], \end{aligned} \quad (17)$$

where we define $\boldsymbol{\beta} = [\cos(\|\mathbf{e}_p[t]\|), \sin(\|\mathbf{e}_p[t]\|)]^*$, $\bar{\eta}_f = \sqrt{1-\eta_f^2}$ and $\mathbf{x}[t] = [\hat{\mathbf{g}}[t-1]^*, \mathbf{e}_p[t]^*]^* \mathbf{z}[t]$ to simplify notation. Since $\mathbf{z}[t] \sim \mathcal{CN}(0, \frac{1}{L} \mathbf{I}_L)$ by Assumption 1, and since the matrix $[\hat{\mathbf{g}}[t-1], \mathbf{e}_p[t]]$ is unitary and independent of $\mathbf{z}[t]$ by definition, the vector $\mathbf{x}[t]$ has distribution $\mathcal{CN}(0, \frac{1}{L} \mathbf{I}_2)$ by the unitary invariance of the Gaussian distribution [37]. Expanding the predictor's objective function, we get

$$\begin{aligned} \mathbb{E} [|\tilde{\mathbf{g}}[t]^* \mathbf{h}[t]|^2] &= \mathbb{E} [\eta_f^2 \mathbf{h}[t-1]^* [\hat{\mathbf{g}}[t-1], \mathbf{e}_p[t]] \boldsymbol{\beta} \boldsymbol{\beta}^* [\hat{\mathbf{g}}[t-1], \mathbf{e}_p[t]]^* \mathbf{h}[t-1] + \\ &\quad 2\text{Re} \{ \eta_f \bar{\eta}_f \mathbf{h}[t-1]^* [\hat{\mathbf{g}}[t-1], \mathbf{e}_p[t]] \boldsymbol{\beta} \boldsymbol{\beta}^* \mathbf{x}[t] \} + \bar{\eta}_f^2 \mathbf{x}[t]^* \boldsymbol{\beta} \boldsymbol{\beta}^* \mathbf{x}[t]] \\ &\stackrel{(a)}{=} \mathbb{E} [\eta_f^2 \mathbf{h}[t-1]^* [\hat{\mathbf{g}}[t-1], \mathbf{e}_p[t]] \boldsymbol{\beta} \boldsymbol{\beta}^* [\hat{\mathbf{g}}[t-1], \mathbf{e}_p[t]]^* \mathbf{h}[t-1] + \bar{\eta}_f^2 \frac{1}{L} \\ &\stackrel{(b)}{=} \eta_f^2 \cos^2(\|\mathbf{e}_p[t]\|) \mathbb{E} [|\mathbf{h}[t-1]^* \hat{\mathbf{g}}[t-1]|^2] + \eta_f^2 \sin^2(\|\mathbf{e}_p[t]\|) \mathbb{E} [|\mathbf{h}[t-1]^* \mathbf{e}_p[t]|^2] \\ &\quad + \mathbb{E} [2\text{Re} \{ \eta_f^2 \cos(\|\mathbf{e}_p[t]\|) \sin(\|\mathbf{e}_p[t]\|) \mathbf{h}[t-1]^* \hat{\mathbf{g}}[t-1] \mathbf{e}_p[t]^* \mathbf{h}[t-1] \}] + \bar{\eta}_f^2 \frac{1}{L}, \end{aligned}$$

where (a) follows because $\mathbf{x}[t]$ is a zero-mean vector and independent of $\mathbf{h}[t-1]^* [\hat{\mathbf{g}}[t-1], \mathbf{e}_p[t]] \boldsymbol{\beta} \boldsymbol{\beta}^*$ and $\mathbb{E} [\mathbf{x}[t]^* \boldsymbol{\beta} \boldsymbol{\beta}^* \mathbf{x}[t]] = \frac{1}{L}$. The expectation of the third term in (b) is zero as a result of Assumption 2 and the fact that $\mathbf{e}_p[t]$ and $\hat{\mathbf{g}}[t-1]$ are instantaneously orthogonal. Moreover, leveraging the independence of $\|\mathbf{h}[t-1]\|$ and $\mathbf{h}[t-1]/\|\mathbf{h}[t-1]\|$, and that $\mathbb{E} [\|\mathbf{h}[t-1]\|^2] = 1$, the objective function simplifies to

$$\mathbb{E} [|\tilde{\mathbf{g}}[t]^* \mathbf{h}[t]|^2] = \eta_f^2 \cos^2(\|\mathbf{e}_p[t]\|) \mathbb{E} [|\mathbf{g}[t-1]^* \hat{\mathbf{g}}[t-1]|^2] + \eta_f^2 \sin^2(\|\mathbf{e}_p[t]\|) \mathbb{E} [|\mathbf{g}[t-1]^* \mathbf{e}_p[t]|^2] + \bar{\eta}_f^2 \frac{1}{L}.$$

Letting $\hat{\rho} = \mathbf{g}[t-1]^* \hat{\mathbf{g}}[t-1]$, the channel $\mathbf{g}[t-1]$ can be written as $\hat{\rho} \hat{\mathbf{g}}[t-1] + \sqrt{1-|\hat{\rho}|^2} \tilde{\mathbf{e}}[t-1]$ where $\tilde{\mathbf{e}}[t-1]$ is a unit norm tangent vector representing the quantization error at time $t-1$. By Assumption 2, $\tilde{\mathbf{e}}[t-1]$ is uniformly distributed in the tangent space and independent of $|\hat{\rho}|$, which results in

$$\begin{aligned} \mathbb{E} [|\tilde{\mathbf{g}}[t]^* \mathbf{h}[t]|^2] &= \eta_f^2 \cos^2(\|\mathbf{e}_p[t]\|) \mathbb{E} [|\hat{\rho}|^2] + \eta_f^2 \sin^2(\|\mathbf{e}_p[t]\|) (1 - \mathbb{E} [|\hat{\rho}|^2]) \mathbb{E} [|\tilde{\mathbf{e}}[t-1]^* \mathbf{e}_p[t]|^2] + \bar{\eta}_f^2 \frac{1}{L} \\ &\stackrel{(a)}{=} \eta_f^2 \cos^2(\|\mathbf{e}_p[t]\|) \mathbb{E} [|\hat{\rho}|^2] + \eta_f^2 \sin^2(\|\mathbf{e}_p[t]\|) (1 - \mathbb{E} [|\hat{\rho}|^2]) \frac{1}{L-1} + \bar{\eta}_f^2 \frac{1}{L}, \end{aligned} \quad (18)$$

where (a) is since the isotropic distribution of $\tilde{\mathbf{e}}[t-1]$ implies that the term $|\tilde{\mathbf{e}}[t-1]^* \mathbf{e}_p[t]|^2$ is beta distributed with mean $\mathbb{E} [|\tilde{\mathbf{e}}^* \mathbf{e}_p[t]|^2] = 1/(L-1)$. Therefore the predictor's objective function is independent of the predicted direction of $\mathbf{e}_p[t]$. Further, $\mathbb{E} [|\hat{\rho}|^2]$ is practically such that $\mathbb{E} [|\hat{\rho}|^2] > (1 - \mathbb{E} [|\hat{\rho}|^2])$, i.e. the quantizer's output is expected to be a more accurate estimate of the channel than a random vector. Thus $\mathbb{E} [|\hat{\rho}|^2] > (1 - \mathbb{E} [|\hat{\rho}|^2])/(L-1)$ and $\mathbb{E} [|\tilde{\mathbf{g}}[t]^* \mathbf{h}[t]|^2]$ is maximized in (18) by letting $\|\mathbf{e}_p[t]\| = 0$ which achieves

$$\mathbb{E} [|\tilde{\mathbf{g}}[t]^* \mathbf{h}[t]|^2] = \eta_f^2 \mathbb{E} [|\hat{\rho}|^2] + \bar{\eta}_f^2 \frac{1}{L}. \quad (19)$$

As a result, the predictor's objective function is maximized by selecting $\|\mathbf{e}_p[t]\| = 0$ and thus $\tilde{\mathbf{g}}[t] = \hat{\mathbf{g}}[t-1]$. This coincides with the differential feedback strategy presented, thus the performance of the proposed framework can not be further improved by such a predictor for the channel model considered.

The derivation in this section, however, does not preclude potential benefits from more sophisticated prediction functions, especially when the band-limited nature of the channel's Doppler spectrum is exploited. While a rich body of literature on channel prediction does exist [38]–[40], work on predicting processes that evolve on the Grassmann manifold is limited to [23], [24] which use predictors and channel models similar to the ones considered in this section. The design of more sophisticated Grassmannian predictors that better exploit channel structure is left for future work.

V. DESIGN CONSIDERATIONS & CODEBOOK CONSTRUCTION

In this section, we discuss the proper initialization of the proposed algorithm and construct adaptive quantization codebooks for the tangent magnitude and direction.

A. Initialization

The proper operation of the Grassmannian differential feedback algorithm is ensured by the fact that, at each iteration, both transmitter and receiver can calculate a common quantized channel based on the quantized tangent vector and the previous quantized channel. For the output of the algorithm to be the same at the transmitter and receiver, however, a common initial vector, $\hat{\mathbf{g}}[0]$, is required. Reinitialization can also be used to recover from feedback bit errors. This initial vector can, for example, be an estimate of $\hat{\mathbf{g}}[0]$ that is fed back from receiver to transmitter in the first iteration along with $\hat{\mathbf{e}}[0]$. Alternatively, it can be based on a memoryless quantization of the channel using methods such as [30], [31]. The additional overhead incurred is amortized when there are many subsequent feedback stages. Another option is to initialize the estimate $\hat{\mathbf{g}}[0]$ to a common random vector from which the algorithm will be able to recover. In this paper, we use memoryless quantization with a random vector codebook at $t = 1$.

B. Tangent Magnitude Quantization

The tangent vector calculated in (14) can be decomposed into a tangent magnitude and a unit norm tangent direction. We propose to quantize the tangent magnitude and direction separately, which ensures that search complexity remains manageable. The problem of quantizing the tangent magnitude is that of quantizing a non-negative scalar. We propose to quantize the tangent magnitude using a Euclidean distance distortion metric

$$\hat{e}_{\text{mag}}[t] = \arg \min_{e_i \in \mathcal{C}_{\text{mag}}} ||\mathbf{e}[t]|| - e_i, \quad (20)$$

where \mathcal{C}_{mag} is the magnitude quantization codebook. The index of the minimizer is then sent to the transmitter via a delay and error-free link that requires $N_{\text{mag}} = \log(|\mathcal{C}_{\text{mag}}|)$ bits. A locally optimal quantization codebook can be found algorithmically using Lloyd's algorithm [41] given either the probability density function (pdf) of the tangent magnitude, or a training set of magnitudes which provide an empirical probability mass function. Finding the exact pdf of the tangent magnitude, however, has been intractable analytically thus far. Generating a good representative training set as input to the Lloyd algorithm is also difficult. The difficulty in both these approaches is that the magnitude pdf depends on the quantization process itself. That is to say that quantizing the tangent direction and magnitude itself could change their distribution in the next step. What is needed is therefore finding a pdf that is a fixed point of the quantization process, which has proven to be difficult. One solution which is adopted in [24] is to uniformly quantize a range of magnitudes, $||\mathbf{e}[t]|| \in [0, 1]$. Unfortunately, [24] has shown that quantization error in the magnitude creates an error floor beyond which CSI can not be improved in highly correlated channels where predictive feedback is most useful. In the limit of perfectly static channels, such magnitude quantization prevents feedback from converging to perfect CSI.

We propose to adapt the quantization range, $[\alpha \hat{e}_{\text{min}}[t], \beta \hat{e}_{\text{max}}[t]]$, to the dynamics of the system and to uniformly quantize the magnitude in that range. To do so, the algorithm maintains a *sliding window* of the quantization range of interest. This adaptive process maintains a running average of the tangent vector magnitude and uses it to update the lower and upper limits of the quantization range. The quantization range, or window, is calculated using quantities that are commonly available at both transmitter and receiver. As a result, both nodes can independently keep track of the current quantization range.

The moving average of quantized tangent magnitudes is given by:

$$\hat{e}_{\text{avg}}[t+1] = \left(1 - \frac{1}{\tau}\right) \hat{e}_{\text{avg}}[t] + \tau \hat{e}_{\text{mag}}[t], \quad (21)$$

where τ is the smoothing factor that can be adjusted by the system. At each iteration either limit of the

quantization range is then updated as follows:

$$\begin{aligned}\hat{e}_{\min}[t+1] &= \begin{cases} (1 - \frac{1}{\tau}) \hat{e}_{\min}[t] + \frac{1}{\tau} \hat{e}_{\text{mag}}[t] & \hat{e}_{\text{mag}}[t] \leq \hat{e}_{\text{avg}}[t] \\ \hat{e}_{\min}[t] & \hat{e}_{\text{mag}}[t] \geq \hat{e}_{\text{avg}}[t] \end{cases} \\ \hat{e}_{\max}[t+1] &= \begin{cases} \hat{e}_{\max}[t] & \hat{e}_{\text{mag}}[t] \leq \hat{e}_{\text{avg}}[t] \\ (1 - \frac{1}{\tau}) \hat{e}_{\max}[t] + \frac{1}{\tau} \hat{e}_{\text{mag}}[t] & \hat{e}_{\text{mag}}[t] \geq \hat{e}_{\text{avg}}[t]. \end{cases}\end{aligned}$$

This allows the feedback algorithm to accurately track the statistics of the tangent magnitude and quantize the current range of interest with higher resolution. In static channels, this allows the system to converge to perfect CSI removing the error floor caused by constant magnitude codebooks.

The fixed scalars α and β must now be such that $\alpha < 1 < \beta$ to enable the flexible adjustment of the window $[\hat{e}_{\min}[t], \hat{e}_{\max}[t]]$. To see this, note that to let $\hat{e}_{\min}[t]$ decrease, it is necessary to be able to quantize values less than $\hat{e}_{\min}[t]$. Similarly, allowing $\hat{e}_{\max}[t]$ to increase requires quantizing values greater than it. In simulation we use $\tau = 5$, $\alpha = 1/2$ and $\beta = 2$. While the chosen parameter values yield good results as shown in Section VII, system performance can be further improved by optimizing over the values of τ , α , and β .

C. Tangent Direction Quantization

The problem of quantizing the tangent direction vector is that of quantizing a unit norm vector which lies in the tangent space orthogonal to the base vector $\hat{\mathbf{g}}[t-1]$, that is $\mathbf{e}[t]^* \hat{\mathbf{g}}[t-1] = 0$. Grassmannian codebooks often used for limited feedback [30], [31] are not suitable for tangent vector quantization since they do not enforce the structural constraint that requires the tangent direction codewords to be orthogonal to the base vector $\hat{\mathbf{g}}[t-1]$. If the quantized tangent vector is not orthogonal to the base vector $\hat{\mathbf{g}}[t-1]$, the geodesic path is undefined and the output of $G(\hat{\mathbf{g}}[t-1], \hat{\mathbf{e}}[t], \ell)$ does not lie on the manifold. Finally, note that the tangent space changes for each base vector $\hat{\mathbf{g}}[t-1]$ and thus the codebook must be adapted [42], i.e., fixed codebooks should not be used.

To respect the varying tangent space geometry and orthogonality constraints, we propose a canonical generating codebook that is adapted at each iteration; only the canonical codebook is actually stored. This canonical codebook quantizes the tangent space at a special $L \times 1$ base vector. The canonical codebook is rotated at each channel realization to match the geometry of the current tangent space.

Definition 3: The N_{dir} -bit canonical tangent codebook is a codebook of $2^{N_{\text{dir}}}$ unit norm vectors orthogonal to the special base vector $x_b := [1, 0, \dots, 0]$. Because of the structure of x_b , the generator codebook, $\mathcal{C}_{\text{dir}}(\mathbf{x}_b)$, has elements, $\mathbf{v}_\ell \in \mathcal{C}_{\text{dir}}(\mathbf{x}_b)$, of the form $\mathbf{v}_\ell = [0, \tilde{\mathbf{v}}_\ell^*]^*$ where $\tilde{\mathbf{v}}_\ell$ is an $(L-1) \times 1$

unit norm vector.

The canonical codebook can be constructed by appending a leading 0 to a codebook of $(L - 1) \times 1$ vectors. The canonical codebook, which can be used to quantize the tangent space at \mathbf{x}_b , can now be rotated at each iteration (or channel realization) to quantize the tangent space at a new vector \mathbf{x} . We define this rotation as follows [23], [42].

Definition 4: Let $\mathbf{U}(\mathbf{x}) : \mathbb{C}^{L \times 1} \mapsto \mathcal{U}^{L \times L}$ be the function that determines the unitary rotation matrix that rotates \mathbf{x}_b to \mathbf{x} , i.e. $\mathbf{U}(\mathbf{x})\mathbf{x}_b = \mathbf{x}$.

Perhaps the easiest way to find such a rotation is to consider the Householder matrix. Let $\mathbf{H}(\mathbf{x}, \mathbf{x}_b) = \mathbf{I}_L - \mathbf{u}\mathbf{u}^*/\mathbf{u}^*\mathbf{x}_b$ where $\mathbf{u} = \mathbf{x}_b - \mathbf{x}$ [43]. Note that $\mathbf{H}(\mathbf{x}, \mathbf{x}_b)$ is a unitary matrix and that if we let $\mathbf{U}(\mathbf{x}) = \mathbf{H}(\mathbf{x}, \mathbf{x}_b)$ then $\mathbf{U}(\mathbf{x})\mathbf{x}_b = \mathbf{x}$ as required. Using the rotation operation it is possible to rotate the canonical codebook to the direction of any particular base point and generate tangent direction codebooks for an arbitrary channel. For a base point, $\hat{\mathbf{g}}[t-1]$, the quantizer chooses $\hat{\mathbf{e}}_{\text{dir}}[t]$ from the rotated codebook

$$\hat{\mathbf{e}}_{\text{dir}}[t] \in \mathcal{C}_{\text{dir}}(\hat{\mathbf{g}}[t-1]) = \{\mathbf{U}(\hat{\mathbf{g}}[t-1])\mathbf{v}_1, \mathbf{U}(\hat{\mathbf{g}}[t-1])\mathbf{v}_2, \dots, \mathbf{U}(\hat{\mathbf{g}}[t-1])\mathbf{v}_{2^{N_{\text{dir}}}}\}.$$

It can be verified that all vectors $\mathbf{U}(\hat{\mathbf{g}}[t-1])\mathbf{v}_\ell$ lie in the tangent plane at $\hat{\mathbf{g}}[t-1]$ since $\hat{\mathbf{g}}[t-1]^*\mathbf{U}(\hat{\mathbf{g}}[t-1])\mathbf{v}_\ell = \mathbf{x}_b^*\mathbf{v}_\ell = 0$, $\forall \ell$.

The proposed codebook design allows us to rotate a canonical codebook to perfectly match the tangent plane at each iteration. This ensures that the output of the Grassmannian differential feedback algorithm remains on the manifold. Moreover, if changes in the channel are isotropic in the L dimensional space, then so are the normalized tangent vectors, implying good performance with canonical codebooks that uniformly quantize the tangent space at \mathbf{x}_b . In Section VII, good performance is achieved even with random vector canonical codebooks. In general, for a given correlation matrix $\mathbf{R}_{\mathbf{h}_{k,\ell}}$, good canonical codebooks can be generated using Lloyd's algorithm, and different codebooks can be stored for environments with different correlation properties.

To formalize the tangent direction quantization, recall that the estimated channel in the next iteration is $G(\hat{\mathbf{g}}[t-1], \hat{\mathbf{e}}[t], 1)$. Given that the loss in sum rate is related to the chordal distance between the channel and its estimate, the quantized tangent direction is computed as

$$\hat{\mathbf{e}}_{\text{dir}}[t][t] = \arg \min_{\mathbf{x}_i \in \mathcal{C}_{\text{dir}}(\hat{\mathbf{g}}[t-1])} d(G(\hat{\mathbf{g}}[t-1], \hat{\mathbf{e}}_{\text{mag}}[t]\mathbf{x}_i, 1), \mathbf{g}[t]) = \arg \max_{\mathbf{x}_i \in \mathcal{C}_{\text{dir}}(\hat{\mathbf{g}}[t-1])} |G(\hat{\mathbf{g}}[t-1], \hat{\mathbf{e}}_{\text{mag}}[t]\mathbf{x}_i, 1)^*\mathbf{g}[t]|^2, \quad (22)$$

where the tangent magnitude, $\hat{\mathbf{e}}_{\text{mag}}[t]$, is given by the output of the magnitude quantization step. We note that although (22) intuitively includes a rotation of the full canonical codebook, computational complexity can be reduced by noticing that an equivalent rotation can be applied to the channels $\hat{\mathbf{g}}[t-1]$ and $\mathbf{g}[t]$

instead, and the codebook need not be actually rotated. The details of this simplification are omitted but can be seen by expanding the objective function $d(G(\hat{\mathbf{g}}[t-1], \hat{e}_{\text{mag}}[t]\mathbf{x}_i, 1)$.

VI. ANALYTICAL PERFORMANCE CHARACTERIZATION

A complete performance analysis of differential or predictive quantization is difficult due to its recursive nature and, in our case, due to the CSI's Grassmannian structure [21]. Relevant performance analysis when the quantized information evolves on the Grassmannian manifold is limited to incomplete performance characterizations [23], [24]. The analysis of the Householder scheme in [23] is limited to characterizing the performance of an open loop predictor, similar to Section IV-B. The resulting lower bound on performance in [23] neglects the quantization process and is not a function of the number of feedback bits. The performance analysis in [24] presents partial distortion bounds that are independent of the channel's temporal correlation and Doppler spread.

In this section we derive an approximation for the distortion achieved by the proposed quantization strategy. Unlike the analysis in [23], [24], our characterization accounts for both the number of feedback bits and the channel's temporal correlation, thus capturing most of the fundamental system parameters. Given the distortion expression, we leverage our analysis in [15] and the discussion in Section III-B to characterize IA's mean loss in sum rate when the proposed feedback strategy is used.

We derive an approximation for the quantizer accuracy defined as $\mathbb{E} [|\mathbf{g}[t]^* \hat{\mathbf{g}}[t]|^2]$, under three mild assumptions. First, we specialize our analysis to the case of channels with relatively low Doppler spread where differential feedback techniques yield the most gain [21], [23]. In the low Doppler setting, we can make use of a small angle approximation [24], [34]. Second, we neglect the error introduced by the tangent magnitude quantizer of Section V. Effectively we make the following assumption.

Assumption 5: We assume perfect magnitude quantization, i.e. $\hat{e}_{\text{mag}}[t] = \|\mathbf{e}[t]\|, \forall t$.

This is a reasonable assumption for our analysis as the numerical results in Section VII-A indicate that the magnitude quantizer adopted achieves very low quantization error with as little as one quantization bit and approaches perfect quantization at all Doppler. Finally, the analysis re-invokes Assumption 1 for simplicity of exposition.

We start by expanding $\mathcal{D}(\eta_f, N_{\text{dir}}) = \mathbb{E} [|\mathbf{g}[t]^* \hat{\mathbf{g}}[t]|^2]$ as

$$\begin{aligned} \mathcal{D}(\eta_f, N_{\text{dir}}) &= \mathbb{E} \left[\left| \left(\cos(\|\mathbf{e}[t]\|) \hat{\mathbf{g}}[t-1] + \sin(\|\mathbf{e}[t]\|) \frac{\mathbf{e}[t]}{\|\mathbf{e}[t]\|} \right)^* (\cos(\hat{e}_{\text{mag}}[t]) \hat{\mathbf{g}}[t-1] + \sin(\hat{e}_{\text{mag}}[t]) \hat{\mathbf{e}}_{\text{dir}}[t]) \right|^2 \right] \\ &= \mathbb{E} \left[\left| \cos(\|\mathbf{e}[t]\|) \cos(\hat{e}_{\text{mag}}[t]) + \sin(\|\mathbf{e}[t]\|) \sin(\hat{e}_{\text{mag}}[t]) \frac{\mathbf{e}[t]^* \hat{\mathbf{e}}_{\text{dir}}[t]}{\|\mathbf{e}[t]\|} \right|^2 \right]. \end{aligned}$$

By invoking Assumption 5, $\mathcal{D}(\eta_f, N_{\text{dir}})$ simplifies to

$$\mathcal{D}(\eta_f, N_{\text{dir}}) = \mathbb{E} \left[\cos^4(\|\mathbf{e}[t]\|) + 2 \cos^2(\|\mathbf{e}[t]\|) \sin^2(\|\mathbf{e}[t]\|) \text{Re} \left\{ \frac{\mathbf{e}[t]^* \hat{\mathbf{e}}_{\text{dir}}[t]}{\|\mathbf{e}[t]\|} \right\} + \sin^4(\|\mathbf{e}[t]\|) \left| \frac{\mathbf{e}[t]^* \hat{\mathbf{e}}_{\text{dir}}[t]}{\|\mathbf{e}[t]\|} \right|^2 \right]. \quad (23)$$

Recall that at each new channel realization the quantized tangent direction $\hat{\mathbf{e}}_{\text{dir}}[t]$ is selected to minimize the chordal distance between $\mathbf{g}[t]$ and $\hat{\mathbf{g}}[t]$, equivalently maximizing $|\mathbf{g}[t]^* \hat{\mathbf{g}}[t]|^2$. By invoking Assumption 5, maximizing $|\mathbf{g}[t]^* \hat{\mathbf{g}}[t]|^2$ becomes equivalent to maximizing the instantaneous value of the right hand side of (23). The distortion function in (23), however, is new; it is neither a Euclidean quantizer's distortion function nor is it a Grassmannian quantizer's phase-invariant distortion function. As a result, neither Euclidean nor Grassmannian quantization results can be used to find an expression for $\mathcal{D}(\eta_f, N_{\text{dir}})$.

By analyzing a similar alternative quantization strategy, however, it is possible to gain insight into the behavior of (23) and derive an approximation for $\mathcal{D}(\eta_f, N_{\text{dir}})$. Instead of directly quantizing the tangent direction to maximize the instantaneous value of (23), we propose to split $\hat{\mathbf{e}}_{\text{dir}}[t]$ into two components, $\hat{\mathbf{e}}_{\text{dir}}[t] = e^{-j\hat{\theta}_e[t]} \hat{\mathbf{e}}_{\mathcal{G}}[t]$. The variables $\hat{\theta}_e[t]$ and $\hat{\mathbf{e}}_{\mathcal{G}}[t]$ are then quantized separately using N_θ and $N_{\mathcal{G}}$ bits. To keep the amount of feedback constant, we require $N_\theta + N_{\mathcal{G}} = N_{\text{dir}}$. Such separation, naturally incurs an increase in distortion when compared to our original “joint” quantization strategy. Therefore, after deriving a distortion expression for the separate quantization strategy, we pick N_θ and $N_{\mathcal{G}}$ to optimize its performance which leads to a more accurate characterization of the original “joint” quantization strategy.

Inserting the proposed tangent decomposition into (23) yields

$$\mathcal{D}(\eta_f, N_{\text{dir}}) = \mathbb{E} \left[\cos^4(\|\mathbf{e}[t]\|) + 2 \cos^2(\|\mathbf{e}[t]\|) \sin^2(\|\mathbf{e}[t]\|) \text{Re} \left\{ \frac{e^{-j\hat{\theta}_e[t]} \mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|} \right\} + \sin^4(\|\mathbf{e}[t]\|) \left| \frac{e^{-j\hat{\theta}_e[t]} \mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|} \right|^2 \right] \quad (24)$$

$$\approx \mathbb{E} \left[\cos^4(\|\mathbf{e}[t]\|) + 2 \cos^2(\|\mathbf{e}[t]\|) \sin^2(\|\mathbf{e}[t]\|) \left| \frac{\mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|} \right| \cos \left(\angle \mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t] - \hat{\theta}_e[t] \right) + \sin^4(\|\mathbf{e}[t]\|) \left| \frac{\mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|} \right|^2 \right], \quad (25)$$

where $\angle \mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]$ is the phase of the inner product $\mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]$ and as a result $\text{Re} \left\{ e^{-j\hat{\theta}_e[t]} \mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t] \right\} = \text{Re} \left\{ e^{j(\angle \mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t] - \hat{\theta}_e[t])} |\mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]| \right\} = |\mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]| \cos \left(\angle \mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t] - \hat{\theta}_e[t] \right)$. Due to the decomposition of $\hat{\mathbf{e}}_{\text{dir}}[t]$ the distortion function decouples, i.e. the quantization of $\hat{\mathbf{e}}_{\text{dir}}[t]$ decouples into a scalar quantization

step to minimize $|\angle \mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t] - \hat{\theta}_e[t]|$ and a Grassmannian quantization step to maximize $|\mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]|$. Thanks to the decomposition, we can now use known results about scalar quantizers and Grassmannian quantizers to analyze the performance of our decomposed strategy.

Recalling Assumption 1 and specializing the analysis to channels with relatively low Doppler as stated earlier, we apply a small angle approximation which can be used to see that $\mathbf{e}[t] \approx \sqrt{1 - \eta_f^2} \mathbf{z}[t]$ [24], [34]. Therefore by Assumption 1 and the small angle approximation, one can see that the tangent direction $\mathbf{e}[t]/\|\mathbf{e}[t]\|$ is isotropic and independent of $\|\mathbf{e}[t]\|$ and that $\sin^2(\|\mathbf{e}[t]\|) = \|\mathbf{e}[t]\|^2 = \|\sqrt{1 - \eta_f^2} \mathbf{z}[t]\|^2$. Using this fact, along with the approximation $\mathbb{E}[x^2] \approx \mathbb{E}[x]^2$, which is accurate when x has relatively low variance, gives

$$\begin{aligned} \mathcal{D}(\eta_f, N_{\text{dir}}) \approx & \eta_f^4 + 2\eta_f^2(1 - \eta_f^2) \mathbb{E} \left[\cos \left(\angle \mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t] - \hat{\theta}_e[t] \right) \right] \mathbb{E} \left[\left| \frac{\mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|} \right| \right] + \\ & (1 - \eta_f^2)^2 \mathbb{E} \left[\left| \frac{\mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|} \right|^2 \right]. \end{aligned} \quad (26)$$

Now, using an N_θ -bit uniform codebook over $[0, 2\pi]$ for $\hat{\theta}_e[t]$ yields

$$\mathbb{E} \left[\cos \left(\angle \mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t] - \hat{\theta}_e[t] \right) \right] = \frac{\sin(\pi/2^{N_\theta})}{\pi/2^{N_\theta}}, \quad (27)$$

since the $\mathbf{e}[t]$ is isotropic which implies that $\angle \mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]$ is uniform [30]. Further, using an $L - 1$ dimensional Grassmannian codebook to construct the canonical codebook for $\hat{\mathbf{e}}_{\mathcal{G}}[t]$ yields

$$\left| \frac{\mathbf{e}[t]^* \hat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|} \right|^2 < 1 - \frac{4}{2^{N_{\mathcal{G}}/(L-2)}}. \quad (28)$$

The proof of (28) follows from the derivations in [31] and is thus omitted. Inserting (27) and (28) into (26) yields the final result which we summarize in the following proposition.

Proposition 6: Under Assumptions 1 and 5, the accuracy achieved by the proposed Grassmannian differential feedback strategy with a $2^{N_{\text{dir}}}$ direction codebook is

$$\begin{aligned} \mathcal{D}(\eta_f, N_{\text{dir}}) \approx & \max_{N_\theta, N_{\mathcal{G}}} \eta_f^4 + 2\eta_f^2(1 - \eta_f^2) \frac{\sin(\pi/2^{N_\theta})}{\pi/2^{N_\theta}} \sqrt{1 - \frac{4}{2^{N_{\mathcal{G}}/(L-2)}}} + (1 - \eta_f^2)^2 \left(1 - \frac{4}{2^{N_{\mathcal{G}}/(L-2)}} \right) \\ \text{s.t. } & N_\theta + N_{\mathcal{G}} = N_{\text{dir}}. \end{aligned}$$

The values of N_θ and $N_{\mathcal{G}}$ can be easily determined by a search over N_{dir} possibilities.

Finally, we note that the approximation in Proposition 6 can now shed light on the performance of IA with the proposed feedback strategy. For example, it can be used in an identical manner as in [15] to evaluate the mean loss in IA sum rate. Since deriving IA's mean loss sum rate parallels our derivation

in [15], and since the main results of [13], [15] have been summarized via (10) and (11), we present the final result on sum rate loss and refer the reader to [15] for more detail. When the differential feedback strategy, with N_{dir} bits for the tangent direction, is used to fulfill IA's CSI requirement in a channel with $\eta_f = J_0(2\pi f_D T_s)$, the mean loss in sum rate is such that

$$\Delta \bar{R}_{\text{sum}} \lesssim \sum_k \frac{d_k}{N_{\text{sc}}} \log_2 \left(1 + \frac{N_{\text{sc}} P}{\sigma_n^2} \sum_{\ell} \frac{\rho_{k,\ell} (d_{\ell} - \delta_{\ell,k})}{d_{\ell}} (1 - \mathcal{D}(\eta_f, N_{\text{dir}})) \right), \quad (29)$$

where $\delta_{\ell,k}$ is the Kronecker delta function. As a result of (29), IA sum rate with differential feedback is such that $\bar{R}_{\text{sum}} \gtrsim R_{\text{sum}}^{\text{ideal}} - \Delta \bar{R}_{\text{sum}}$, where $R_{\text{sum}}^{\text{ideal}}$ is the sum rate achieved with perfect CSI. We note, however, that the characterization given in (29) and derived using the methods [15] is known to be rather loose especially when CSI quality is poor, mainly due to the use of Jensen's inequality. Tighter, yet more involved, sum rate characterizations can be derived using alternative methods as in [44].

VII. SIMULATION RESULTS

In this section we present numerical performance results. First, we characterize the performance of Grassmannian differential feedback in relation to Doppler and codebook size, and explore the trade-off between codebook size and refresh rate. Second, we demonstrate the performance of IA when CSI is obtained via the Grassmannian differential feedback strategy.

A. Quality of CSI

First, we examine the quantization error resulting from using the proposed algorithm to quantize the channel's impulse response. We consider L tap channels which are temporally correlated according to the M -order autoregressive channel model defined in [36]. Unless otherwise stated, our simulations use $M = 200$ and initialize the time series as in [36]. Although the analysis of Section IV specialized $M = 1$, adopting a more complex higher order model in simulation provides more realistic performance results. We restrict our attention to an uncorrelated uniform fading profile, that is $\mathbf{R}_{\mathbf{h}} = \frac{1}{L} \mathbf{I}_L$. The normalized channel vectors, $\frac{\mathbf{h}_{k,\ell}[t]}{\|\mathbf{h}_{k,\ell}[t]\|}$ constitute the correlated Grassmannian time series to be quantized.

Fig. 3 shows the chordal distance between the actual and quantized channel for a three tap correlated channel with $f_D T_s = 0.005$. Comparing Grassmannian differential feedback to memoryless random vector quantization, we see that the proposed algorithm achieves a much lower chordal distance as early as the third or fourth iteration. The initial transient phase in which the quantization error is higher than the steady state is due to initializing the algorithm with an all ones vector, i.e. $\hat{\mathbf{g}}[0] = \frac{1}{\sqrt{3}} [1, 1, 1]^*$. The quick convergence to the steady state performance indicates that there is limited loss due to random

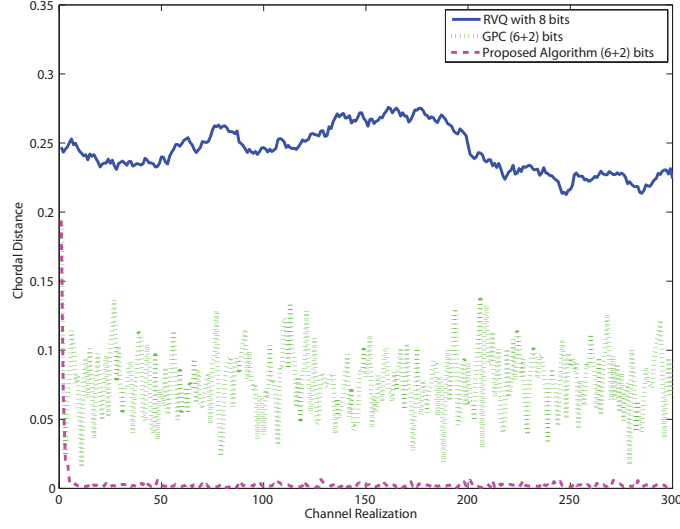


Fig. 3. Chordal distance, $d(\hat{\mathbf{g}}[t], \mathbf{g}[t])$, plotted over time for a slowly channel with normalized Doppler $f_D T_s = 0.003$. This shows the high quality CSI achieved by the proposed algorithm with 6 and 2 bits for channel direction and magnitude respectively, compared to [24] and memoryless quantization with a random codebook of 8 bits.

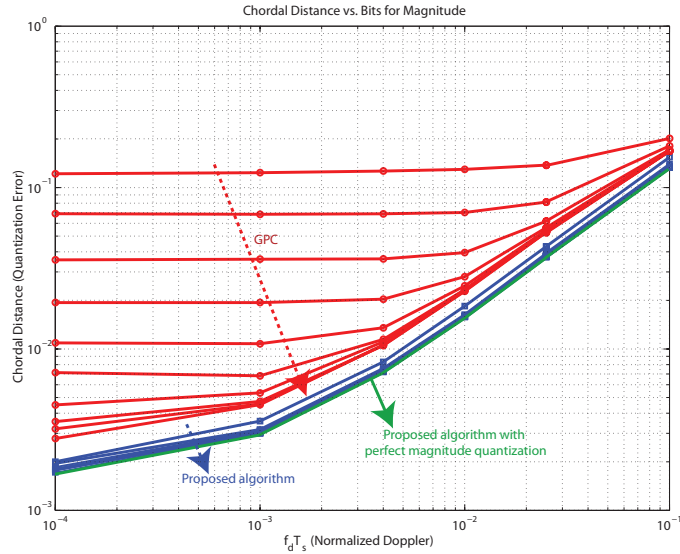


Fig. 4. This shows the average chordal distance, or quantization error, versus the number of bits allocated for the tangent magnitude (1 to 10 bits), while keeping direction bits fixed. This shows the good performance of the proposed algorithm when compared with [24].

initialization. The proposed algorithm also achieves a much lower steady state error compared to the GPC algorithm presented in [24] due to the adaptive codebooks presented. Examining the average quantization error achieved by both the proposed algorithm and the GPC algorithm of [24], we see that the proposed algorithm consistently outperforms. This can be seen in Figs. 4 and 5.

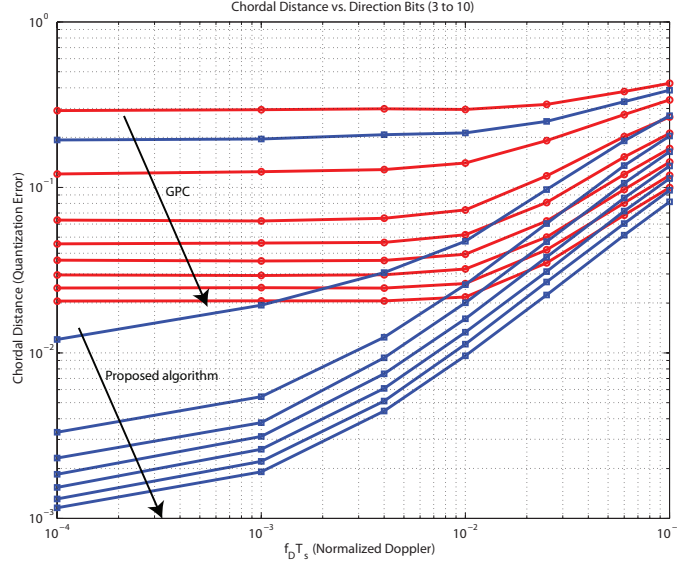


Fig. 5. This shows the average chordal distance, or quantization error, versus the number of bits allocated for the tangent direction, while keeping magnitude bits fixed. This shows the good performance of the proposed algorithm when compared with [24].

Fig. 4 plots the average error of both algorithms vs. Doppler spread for a varying number of magnitude quantization bits, N_{mag} . The adaptive magnitude codebook, with even 1 bit of feedback, achieves 1.5dB lower distortion than the GPC algorithm with 10 bits of magnitude feedback. Adapting the codebook as presented in Section V allows accurate quantization and leaves very little to be gained from increasing the number of magnitude bits. Therefore, while [24] requires an increasing number of magnitude bits at low Doppler, which is the case of interest, our proposed algorithm works well with 1 bit for all Doppler spread. Moreover, if the total number of feedback bits is fixed, and the bits are allocated optimally across the direction and magnitude codebooks, it can be numerically shown that the optimal number of bits allocated to the magnitude is always 1.

Fig. 5 plots the average error for a varying number of direction quantization bits, N_{dir} , while keeping N_{mag} constant. Fig. 5 again shows that the proposed algorithm consistently outperforms existing feedback schemes. We can see that the proposed algorithm can achieve up to 13dB lower quantization error than previous schemes. This is due to the proposed magnitude and tangent codebooks which adapt to the geometry of each point on the manifold. Moreover, Fig. 5 shows that, for the proposed algorithm, increasing the codebook size from 3 to 4, or 4 to 5 bits offers the greatest improvements in quantization error, after which only marginal improvements are experienced. Interestingly, as opposed to further increases in codebook size, going from 3 to 5 bits does not introduce a constant improvement at all

TABLE I
MINIMUM DIRECTION BITS

Length of Channel	Suggested Direction Bits
2	3
3	5
4	7
5	9
6	11

Doppler and instead changes the slope of the curve providing much larger gains in slow fading channels.

Investigating this phase transition behavior further we notice that an error floor exists at low Doppler for all tangent direction codebooks smaller than 5 bits. Moreover, this phase transition is independent of the number of bits allocated to the magnitude, i.e. quantizing the magnitude with infinite precision does not change this behavior. Examining channels with more taps, we notice that this behavior is recurring, and that the phase transition occurs according to Table I. It is interesting to note that this phase transition behavior for an L -tap channel always occurs at $2L - 1$ bits, which is the number of free variables in the tangent direction. A formal investigation of this is left for future work.

Fig. 6 plots $\mathbb{E} [|\mathbf{g}[t]^* \hat{\mathbf{g}}[t]|^2]$ vs. $f_D T_s$ for a practical range of N_{dir} to evaluate the accuracy of the performance characterization derived in Section VI. Similarly to Section VI, we consider the performance of the differential algorithm for a first order auto regressive channel model. As expected, the performance characterization is most accurate for slow fading channels and is reasonably accurate in faster fading scenarios. The reason for this is that the small angle approximation used is naturally suited for relatively low Doppler spreads. Moreover, the performance characterization gradually becomes more accurate as N_{dir} increases. Recall that the result was derived using a “separate” quantization scheme in which the tangent direction was decomposed into a phase angle, and a phase invariant direction. The bit allocation between the 2 variables, N_θ and $N_{\mathcal{G}}$, was then optimized to better characterize the superior “joint quantization” approach. The penalty due to such separate quantization is large when a small number of bits is available as the “product” of the two codebooks does not efficiently fill the tangent space. As the number of bits increases, the penalty due to such separate quantization diminishes and the performance characterization becomes more accurate even at high Doppler. Such behavior is in line with previous results on separate vs. joint quantization [41].

Finally, we highlight an issue that has often been overlooked in CSI feedback, with a few exceptions such as the analysis in [25], [45] for single-user systems. Namely, the numerical analysis thus far has

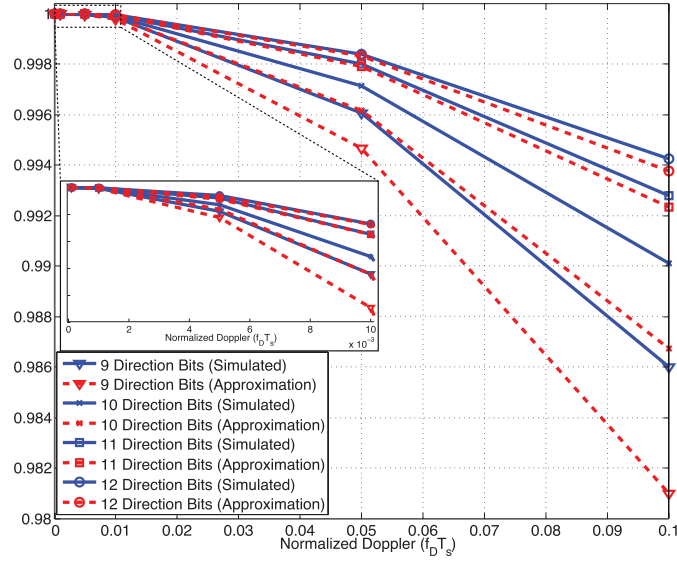


Fig. 6. This figure compares the simulated performance of differential feedback with the performance characterization derived in Section VI. While the characterization is loose when feedback bits are low, due to the significant loss from separate quantization, the characterization becomes increasingly accurate as resolution increases.

assumed a fixed normalized Doppler spread $f_D T_s$, and a fixed number of feedback bits. Both these quantities, however, can be controlled by the system while keeping the overall overhead of feedback, i.e. feedback bit rate, fixed. When such optimization is possible, feedback frequency and codebook size can be traded-off depending on channel conditions, resulting in lower distortion. Fig. 7 shows that increasing codebook size while keeping overall overhead constant can significantly deteriorate performance in fast fading channels. On the one hand, slow fading channels can support a large delay between high resolution feedback updates, since the channel does not change appreciably. On the other hand, decreasing the feedback time T_s to counter the high Doppler frequency f_D yields better performance even if each update carries fewer bits. A formal analysis of this effect in the context of Grassmannian differential feedback is left for future work.

B. IA Performance Analysis

In this section we present simulation results to demonstrate the performance of IA when channel knowledge at the transmitter is obtained via the Grassmannian differential feedback strategy detailed in Section IV. We assume that the channels are correlated according to the same autoregressive model used in Section VII-A, and that all the channels have equal Doppler spreads. Moreover, we assume that $\rho_{k,\ell} = \rho$, $\forall k, \ell$. Thus we define the signal-to-noise ratio as $SNR = \frac{\rho P}{\sigma_n^2}$. Since the frequency extended

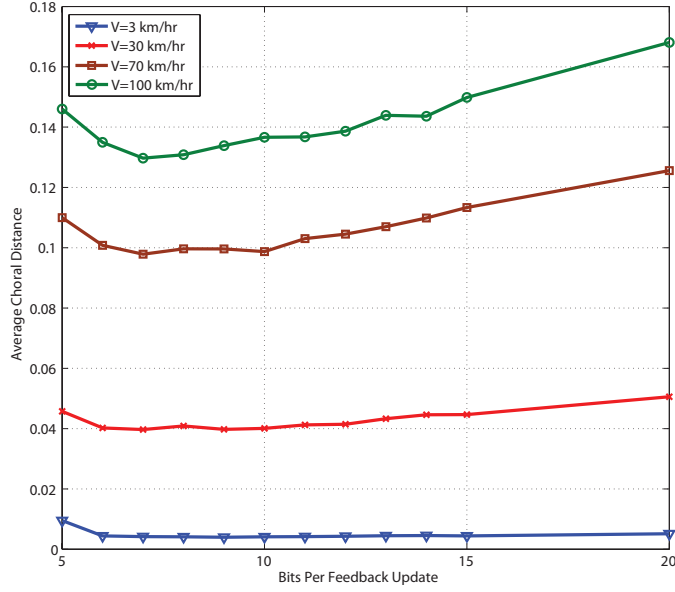


Fig. 7. Keeping the total feedback rate fixed at 5000 bits/s, this figure shows the average chordal distance or quantization error, versus the number of feedback bits. This shows that it is most often optimal to send frequent low resolution feedback. While slow fading channels can support and benefit from large feedback update periods, the performance in faster fading channels significantly deteriorates if refresh rate is sacrificed for codebook size. In all cases $N_{\text{mag}} = 1$ and N_{dir} is varied.

system can be viewed as a virtual MIMO system, the sum rate achieved is calculated as [46], [47],

$$\bar{R}_{\text{sum}} = \sum_{k=1}^K \frac{1}{N_{\text{sc}}} \log_2 \left| \mathbf{I} + \frac{N_{\text{sc}} P}{d_k} \left(\sigma^2 \mathbf{I} + \frac{N_{\text{sc}} P}{d_k} \sum_{m \neq k} \mathbf{H}_{k,m} \hat{\mathbf{F}}_m \hat{\mathbf{F}}_m^* \mathbf{H}_{k,m}^* \right)^{-1} \left(\mathbf{H}_{k,k} \hat{\mathbf{F}}_k \hat{\mathbf{F}}_k^* \mathbf{H}_{k,k}^* \right) \right|, \quad (30)$$

and the precoders, $\hat{\mathbf{F}}_k = [\hat{\mathbf{f}}_k^1, \hat{\mathbf{f}}_k^2, \dots, \hat{\mathbf{f}}_k^{d_k}]$, are calculated given ideal or estimated CSI. We note here that the sum rate calculated in (30) assumes that receivers use the perfect CSI available to them for decoding. For the results in this section, we use a modified version of the interference alignment algorithm proposed in [2], with the number of data symbols given by the conditions in [1]. Although a closed form solution for the IA precoders exists for the single antenna frequency extended interference channel in [1], it has been shown that the solution in [1] yields low sum rates if not further improved as in [48].

Fig. 8 shows the sum rate achieved by 3 users using IA over 16 channel extensions. CSI is obtained by our proposed algorithm, the strategy in [24], and memoryless vector quantization. For all the feedback strategies shown, each 3-tap channel is quantized using 10 bits. For our feedback strategy and that of [24], 7 and 3 bits are used for the tangent direction and magnitude respectively. From Fig. 8, one can show that, in the case of perfect CSI, the rate of increase of sum rate with SNR is 1.35, which approaches the 1.46 degrees of freedom predicted in theory over 16 subcarriers.

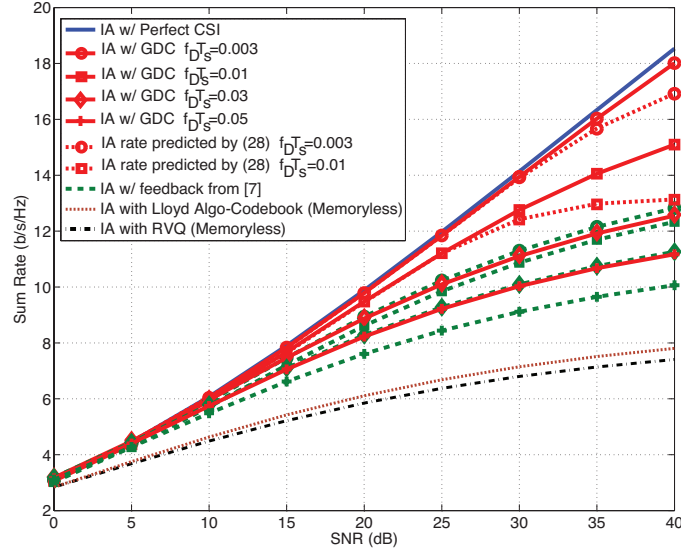


Fig. 8. This figure shows the performance of IA with imperfect CSI obtained through the proposed algorithm, the algorithm in [24], as well as random vector quantization. This shows that for slowly varying channels, the proposed algorithm allows interference alignment networks to come very close to the perfect CSI upper bound.

As for the performance of IA with Grassmannian differential feedback, Fig. 8 shows that the sum rate achieved by IA is significantly improved by exploiting temporal correlation in the channel. Although multiplexing gain will likely never be preserved with a constant number of feedback bits, we see that for correlated channels, the proposed algorithm exhibits close to perfect performance over a wide range of SNR. Practically, this is in fact what is needed. Since a system will almost never function in the asymptotically high SNR regime, the goal in practical systems is to optimize for the medium to high SNR regime where interference alignment is likely to be used. The proposed algorithm succeeds in providing negligible sum rate loss at up to 30dB, in channels with a normalized Doppler of up to 10^{-2} . In fact this can be achieved with much less feedback bits as opposed to the 10 bit feedback performance shown in Fig. 8. The proposed algorithm continues to outperform memoryless quantization even at a significant Doppler of $f_D T_s = 0.05$ regardless of whether random codebooks (RVQ) or optimized codebooks, generated using the well-known Lloyd algorithm [41], [49], are used. The proposed algorithm also consistently outperforms the GPC algorithm in [24] as a result of the adaptive magnitude and direction codebooks used. Finally, Fig. 8 shows the analytical sum rate predicted by (29) for the case of $f_D T_s = 0.003$ and 0.01. As stated in Section VI, the rate characterization is seen to be rather accurate for high SNRs up to 30dB, when 10 bits provide CSI of sufficient accuracy, but becomes progressively looser as SNR increases further and the CSI requirement becomes more stringent.

VIII. CONCLUSION

Limited feedback is a practical way to meet the CSI requirements of IA. To provide high resolution, we proposed a feedback strategy based on Grassmannian differential coding that uses the channels temporal correlation to track channel responses by moving along geodesic paths defined via quantized tangent feedback. We showed that IA in conjunction with the proposed algorithm can benefit from more accurate channel knowledge and provide better sum rate performance for a large range of SNR conditions and a limited number of feedback bits. Moreover, our algorithm exhibits significant improvements over earlier methods of feeding back correlated time series on the manifold due to the optimized tangent and magnitude codebooks.

REFERENCES

- [1] V. Cadambe and S. Jafar, "Interference Alignment and Degrees of Freedom of the K-User Interference Channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.
- [2] S. Peters and R. W. Heath, Jr., "Cooperative algorithms for MIMO interference channels," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 206–218, Jan. 2011.
- [3] K. Gomadam, V. Cadambe, and S. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3309–3322, 2011.
- [4] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, "Minimum mean squared error interference alignment," *Conference Record of the 43rd Asilomar Conference on Signals, Systems and Computers*, pp. 1106–1110, Nov. 2009.
- [5] C. Shi, R. A. Berry, and M. L. Honig, "Adaptive Beamforming in Interference Networks via Bi-Directional Training," *ArXiv preprint arXiv:1003.4764*, Mar. 2010. [Online]. Available: <http://arxiv.org/abs/1003.4764>
- [6] D. Papailiopoulos and A. Dimakis, "Interference alignment as a rank constrained rank minimization," in *Proc. of IEEE Global Telecommunications Conference*, pp. 1–6, Dec. 2010.
- [7] S. W. Choi, S. Jafar, and S.-Y. Chung, "On the beamforming design for efficient interference alignment," *IEEE Communications Letters*, vol. 13, no. 11, pp. 847–849, Nov. 2009.
- [8] I. Santamaría, O. González, R. W. Heath, Jr., and S. W. Peters, "Maximum sum-rate interference alignment algorithms for MIMO channels," in *Proc. of IEEE Global Telecommunications Conference*, pp. 1–6, Dec. 2010.
- [9] R. Tresch, M. Guillaud, and E. Riegler, "On the achievability of interference alignment in the K-user constant MIMO interference channel," in *Proc. of IEEE/SP 15th Workshop on Statistical Signal Processing*, pp. 277–280, Sep. 2009.
- [10] D. Love, R. W. Heath, Jr., V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.
- [11] S. Peters and R. W. Heath, Jr., "Orthogonalization to reduce overhead in MIMO interference channels," in *Proc. of International Zurich Seminar*, pp. 126–129, March 2010.
- [12] M. Guillaud and D. Gesbert, "Interference alignment in the partially connected K-user MIMO interference channel," *Proc. of European Signal Processing Conference (EUSIPCO), Barcelona, Spain*, pp. 1–5, Sept. 2011.
- [13] H. Bolcskei and I. Thukral, "Interference alignment with limited feedback," in *Proc. of IEEE International Symposium on Information Theory*, pp. 1759–1763, Jul. 2009.

- [14] R. Krishnamachari and M. Varanasi, "Interference alignment under limited feedback for MIMO interference channels," in *Proc. of IEEE International Symposium on Information Theory*, pp. 619–623, Jun. 2010.
- [15] O. El Ayach and R. W. Heath, Jr., "Interference alignment with analog channel state feedback," *IEEE Transactions on Wireless Communication*, vol. 11, no. 2, pp. 626–636, Feb. 2012.
- [16] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.
- [17] J. Lee, J. Han, and J. Zhang, "MIMO technologies in 3GPP LTE and LTE-advanced," *EURASIP Journal on Wireless Communications and Networking*, pp. 1–10, March 2009.
- [18] M. Trivellato, F. Boccardi, and H. Huang, "On transceiver design and channel quantization for downlink multiuser mimo systems with limited feedback," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1494–1504, 2008.
- [19] D. Sacristan-Murga and A. Pascual-Iserte, "Differential feedback of MIMO channel gram matrices based on geodesic curves," *IEEE Transactions on Wireless Communications*, vol. 9, no. 12, pp. 3714–3727, Dec. 2010.
- [20] D. Sacristan-Murga, F. Kaltenberger, A. Pascual-Iserte, and A. Perez-Neira, "Differential feedback in MIMO communications: Performance with delay and real channel measurements," in *Proc. of Workshop on Smart Antennas*, pp. 1–8, 2009.
- [21] K. Kim, I. H. Kim, and D. Love, "Utilizing temporal correlation in multiuser MIMO feedback," in *Proc. of 42nd Asilomar Conference on Signals, Systems and Computers*, pp. 121–125, Oct. 2008.
- [22] J. Roh and B. Rao, "Efficient feedback methods for MIMO channels based on parameterization," *IEEE Transactions on Wireless Communications*, vol. 6, no. 1, pp. 282–292, Jan. 2007.
- [23] L. Liu and H. Jafarkhani, "Novel transmit beamforming schemes for time-selective fading multiantenna systems," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4767–4781, Dec. 2006.
- [24] T. Inoue and R. W. Heath, Jr., "Grassmannian predictive coding for limited feedback in multiple antenna wireless systems," *submitted to IEEE Transactions on Signal Processing*, 2011. [Online]. Available: <http://arxiv.org/abs/1105.5782>
- [25] T. Kim, D. Love, and B. Clerckx, "Does frequent low resolution feedback outperform infrequent high resolution feedback for multiple antenna beamforming systems?" *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1654–1669, Apr. 2011.
- [26] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell System Technical Journal*, vol. 47, no. 6, pp. 957–1000, 1968.
- [27] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell mimo cooperative networks: A new look at interference," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, 2010.
- [28] E. Martinian, "Waterfilling gains $O(1/\text{SNR})$ at high SNR." [Online]. Available: <http://www.csua.berkeley.edu/~emin/research/wfill.pdf>.
- [29] B. Nosrat-Makouei, J. Andrews, and R. W. Heath, Jr., "A simple SINR characterization for linear interference alignment over uncertain MIMO channels," in *Proc. of IEEE Int. Symposium on Inf. Theory*, pp. 2288–2292, Jun. 2010.
- [30] D. Love and R. W. Heath, Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.
- [31] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.
- [32] C.-B. Chae, D. Mazzarese, T. Inoue, and R. W. Heath, Jr., "Coordinated Beamforming for the Multiuser MIMO Broadcast Channel With Limited Feedforward," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 6044–6056, Dec. 2008.

- [33] A. Ashikhmin and R. Gopalan, "Grassmannian packings for efficient quantization in MIMO broadcast systems," in *Proc. of IEEE International Symposium on Information Theory*, pp. 1811–1815, Jun. 2007.
- [34] T. Arias, A. Edelman, and S. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, pp. 303–353, 1998.
- [35] R. Bhagavatula, R. W. Heath, Jr., and B. Rao, "Limited feedback with joint CSI quantization for multicell cooperative generalized eigenvector beamforming," in *Proc. of IEEE International Conference on Acoustics Speech and Signal Processing*, pp. 2838–2841, Mar. 2010.
- [36] K. Baddour and N. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1650–1662, Jul. 2005.
- [37] A. Tulino and S. Verdú, *Random matrix theory and wireless communications*. Now Publishers Inc, 2004.
- [38] C. Komninakis, C. Fragouli, A. Sayed, and R. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1065–1076, May 2002.
- [39] Y. Kho and D. Taylor, "MIMO channel estimation and tracking based on polynomial prediction with application to equalization," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1585–1595, May 2008.
- [40] T. Svantesson and A. Swindlehurst, "A performance bound for prediction of MIMO channels," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 520–529, Feb. 2006.
- [41] R. Gray, "Vector quantization," *IEEE ASSP Magazine*, vol. 1, no. 2, pp. 4–29, April 1984.
- [42] R. W. Heath, Jr., T. Wu, and A. Soong, "Progressive refinement for high resolution limited feedback multiuser MIMO beamforming," in *Proc. of 42nd Asilomar Conference on Signals, Systems and Computers*, pp. 743–747, Oct. 2008.
- [43] K.-L. Chung and W.-M. Yan, "The complex Householder transform," *IEEE Transactions on Signal Processing*, vol. 45, no. 9, pp. 2374–2376, Sep. 1997.
- [44] O. El Ayach, A. Lozano, and R. W. Heath, Jr., "On the overhead of interference alignment: Training, feedback, and cooperation," *submitted to IEEE Transactions on Wireless Communication*, Apr. 2012. [Online]. Available: <http://arxiv.org/abs/1204.6100>
- [45] K. Huang, R. Heath, and J. Andrews, "Limited feedback beamforming over temporally-correlated channels," *Signal Processing, IEEE Transactions on*, vol. 57, no. 5, pp. 1959–1975, 2009.
- [46] A. Lozano and A. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Transactions on Information Theory*, vol. 48, no. 12, pp. 3117–3128, 2002.
- [47] R. Blum, "MIMO capacity with interference," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 793–801, Jun. 2003.
- [48] D. Kim and M. Torlak, "Optimization of interference alignment beamforming vectors," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1425–1434, Dec. 2010.
- [49] P. Xia, S. Zhou, and G. Giannakis, "Achieving the welch bound with difference sets," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1900–1907, 2005.